



Identification of Signatures of Selection by Whole-Genome Resequencing of a Chinese Native Pig

Wei Zhang^{1†}, Min Yang^{2†}, Mei Zhou¹, Yuanlang Wang², Xudong Wu², Xiaodong Zhang², Yueyun Ding², Guiying Zhao³, Zongjun Yin^{2*} and Chonglong Wang^{1*}

¹ Key Laboratory of Pig Molecular Quantitative Genetics of Anhui Academy of Agricultural Sciences, Anhui Provincial Key Laboratory of Livestock and Poultry Product Safety Engineering, Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei, China, ² College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, ³ Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming, China

OPEN ACCESS

Edited by:

Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil

Reviewed by:

Samuele Bovo,
University of Bologna, Italy
Jian Xu,
Chinese Academy of Fishery
Sciences (CAFS), China

*Correspondence:

Zongjun Yin
yinzongjun@ahau.edu.cn
Chonglong Wang
ahwchl@163.com

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 27 May 2020

Accepted: 20 August 2020

Published: 17 September 2020

Citation:

Zhang W, Yang M, Zhou M,
Wang Y, Wu X, Zhang X, Ding Y,
Zhao G, Yin Z and Wang C (2020)
Identification of Signatures
of Selection by Whole-Genome
Resequencing of a Chinese Native
Pig. *Front. Genet.* 11:566255.
doi: 10.3389/fgene.2020.566255

Identification of genomic signatures of selection that help reveal genetic mechanisms underlying traits in domesticated pigs is of importance. Anqing six-end-white pig (ASP), a representative of the native breeds in China, has many distinguishing phenotypic characteristics. To identify the genomic signatures of selection of the ASP, whole-genome sequencing of 20 ASPs produced 469.01 Gb of sequence data and more than 26 million single-nucleotide polymorphisms. Combining these data with the available whole genomes of 13 Chinese wild boars, 157 selected regions harboring 48 protein-coding genes were identified by applying the polymorphism levels ($\theta\pi$) and genetic differentiation (F_{ST}) based cross approaches. The genes found to be positively selected in ASP are involved in crucial biological processes such as coat color (*MC1R*), salivary secretion (*STATH*), reproduction (*SPIRE2*, *OSBP2*, *LIMK2*, *FANCA*, and *CABS1*), olfactory transduction (*OR5K4*), and growth (*NPY1R*, *NPY5R*, and *SELENOM*). Our research increased the knowledge of ASP phenotype-related genes and help to improve our understanding of the underlying biological mechanisms and provide valuable genetic resources that enable effective use of pigs in agricultural production.

Keywords: Anqing six-end-white pig, SNP, signatures of selection, F_{ST} , $\theta\pi$

INTRODUCTION

Until the Neolithic Age about 10,000 years ago, humans have changed from nomadic to settled, which make captive breeding possible, and the wild boars were successfully domesticated by human at multiple locations around the world (Giuffra et al., 2000; Kijas and Andersson, 2001; Larson, 2005). Selective breeding has generated approximately 300 pig breeds that are adapted to various environmental conditions and production system (Veirano Fréchu, 2007). Natural and artificial selection played a key role in shaping the fitness of domestic pig to those environments and demands, with the underlying mechanisms being of great interest in evolutionary biology, including the relationship between molecular and phenotypic changes and how the involvement of natural and artificial processes in the evolutionary process have shaped the modern animal genomes.

According to the theory of population genetics, the functional genes subject to selection would reveal characteristic patterns due to selection preference, and these patterns are known as “signatures of selection” (Fan et al., 2014). The selected region is often a chromosomal region with low genetic diversity within the group and high genetic differentiation rate between groups. A series of statistical approaches, based on the genetic diversity and genetic differentiation, have been proposed for the detection of selection signatures, such as genetic differentiation (F_{ST}) and polymorphism levels ($\theta\pi$, pairwise nucleotide variation as a measure of variability). The fixation index (F_{ST}) statistic, which is based on population differentiation, was first defined by Lewontin and Krakauer (1973) based on coefficient F (Wright, 1949) and developed by Weir and Cockerham (1984), Akey et al. (2002), and Gianola et al. (2010). The polymorphism levels ($\theta\pi$) statistic, which is based on genetic diversity, was measured for each individual by nucleotide diversity π (Nei and Li, 1979) and Watterson’s estimator θ (Watterson, 1975) and usually used to identify regions of selection between domesticated and wild species. Surveying the genomic regions with $\theta\pi$ and F_{ST} based cross approaches is useful for the detection of selection signatures (Li et al., 2013).

Recent whole genome-wide scans in diverse pig breeds aimed to uncover the underlying mechanism for complex phenotype–genotype association showed that detecting comprehensive signatures of intense artificial selection could acknowledge the characteristics of well-defined breeds (Groenen et al., 2012; Rubin et al., 2012; Li et al., 2013; Frantz et al., 2015). For example, the selected genes *NR6A1*, *PLAG1*, and *LOCRL* have played a key role on elongation of body length in European domestic pigs. Fang et al. (2009) elucidated that coat color variation is the result of intentional selection at the porcine melanocortin receptor 1 (*MC1R*) locus (Fang et al., 2009). A study identified genomic selection regions at *HIFA*, a master regulator of oxygen homeostasis, in Tibetan wild boar (Li et al., 2013). Ai et al. (2015) and Chen et al. (2018) identified that copy number variation in the *MSRB3* could regulate the size of porcine ear and identified that the *TGFB3* and *DAB2IP* played an important role in regulating the number of ribs (Zhu et al., 2015; Chen et al., 2018). Although many studies have been done in various pig breeds, in view of the diverse phenotypes among the 300 pig breeds, efforts still need to be done in elucidating the phenotype differences resulting from different environments and artificial selection.

Anqing six-end-white pig (ASP) is a representative Chinese indigenous, disease-resistant breed with high fertility, high fat content, excellent meat quality, good maternal stability, and a crude-feed tolerance that has been bred with artificial selection for a long time. In our previous studies, the average backfat thickness and the intramuscular fat content in the longissimus dorsi muscle of ASP at about 100 kg were measured, and the value were 46.01 ± 2.55 , 6.54 ± 0.81 (mean \pm SD, $n = 6$) separately (Wang et al., 2020). Based on molecular and multiomics level, we have identified some genes, microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) played an

important role in meat quality, lipid metabolism, and fat deposition (Zhang et al., 2015; Hu et al., 2019; Ding et al., 2020; Wang et al., 2020). However, the genetics basis of the characteristics in ASP, particularly at the genomic level, remains largely unknown.

To access a comprehensive analysis of genetic variations underlying domestication traits in ASP breed, we used whole-genome resequencing data of 20 unrelated ASP, together with 13 publicly available Asian wild boar genomes, using the F_{ST} and $\theta\pi$ based cross approaches to explore the signatures of selection in ASP. In this study, we identified a suite of genes having undergone positive selection that may contribute to domestication phenotypes, including disease resistance, reproduction, digestive system, and lipid metabolism. The findings herein will provide insights to increase understanding of the genetic basis that determines the unique traits of ASP and provide scientific foundation for its development and utilization.

MATERIALS AND METHODS

Sample Collection, DNA Extraction, and Sequencing

We sampled 20 unrelated ASP (Figure 1), collecting from the ASP conservation farm (Anqing, China; longitude, 116°33'E; latitude, 30°19'N). Genomic DNA was extracted from the ear samples using a standard phenol–chloroform method (Sambrook and Russell, 2001) and was stored at 4°C to avoid freeze–thawing and tested for concentration (as ng/ μ l) using a Nanodrop. The DNA was fragmented and treated following the Illumina DNA sample preparation protocol, with a process of end-repaired, A-tailed, ligated to paired-end adaptors and PCR amplification with 350-bp inserts. The constructed libraries were sequenced on the Illumina HiSeq X Ten platform (Illumina, San Diego, CA) for 150-bp paired-end reads at Novogene (Beijing, China).



FIGURE 1 | Anqing six-end-white pig used in this study.

TABLE 1 | Summary statistic of resequencing data.

Sample	Raw reads	Effective rate (%)	Raw base (G)	Clean base (G)	Coverage	Coverage at least 1 or 4 × (%)
S19	83,823,171	99.7	25.15	25.07	9.64	98.90 or 95.33
S9	86,473,158	99.85	25.94	25.90	9.96	99.02 or 94.58
S18	72,934,572	99.72	21.88	21.82	8.39	98.69 or 92.56
S1	75,402,685	99.79	22.62	22.57	8.68	99.88 or 92.05
S13	82,954,876	99.71	24.89	24.81	9.54	98.72 or 94.99
S12	79,354,645	99.57	23.81	23.70	9.12	98.71 or 93.83
S15	76,047,672	99.74	22.81	22.75	8.75	98.67 or 93.36
S14	88,777,383	99.73	26.63	26.56	10.22	98.80 or 96.13
S17	81,000,552	99.71	24.30	24.23	9.32	98.74 or 94.64
S16	82,788,651	99.73	24.84	24.77	9.53	98.76 or 95.00
S10	77,587,033	99.75	23.28	23.22	8.93	99.01 or 93.02
S20	83,019,325	99.71	24.91	24.83	9.55	98.68 or 94.91
S7	71,153,986	99.84	21.35	21.31	8.20	98.83 or 90.59
S5	71,392,366	99.83	21.42	21.38	8.22	98.78 or 90.32
S11	75,019,716	99.72	22.50	22.43	8.63	98.62 or 92.79
S8	74,686,600	99.84	22.40	22.37	8.60	98.85 or 91.78
S6	75,178,013	99.81	22.55	22.51	8.66	98.95 or 92.17
S4	75,380,120	99.83	22.61	22.57	8.68	98.94 or 92.05
S2	75,927,310	99.82	22.78	22.73	8.74	98.90 or 92.16
S3	74,458,919	99.83	22.34	22.30	8.58	98.91 or 91.96

The raw resequencing reads were filtered using NGSQCToolkit, which removed reads containing adapter or poly-N, low-quality reads with >30% base having Phred quality ≤ 20 , the 5' and 3' ends 5 bp low-quality base of a read.

To comprehensively investigate the signatures of selection of ASP population during domestication and breeding, we also retrieved 13 resequencing data of Asian wild boars (~13× coverage per individual) using as reference group (**Supplementary Table S1**) (Groenen et al., 2012; Li et al., 2013; Ai et al., 2015).

Variant Calling and Annotation

The filtered resequencing reads from all genome were then mapped independently to pig reference genome using BWA-MEM version 0.7.10 (Li and Durbin, 2009) with default parameters. We also downloaded the reads file for 13 pig individuals from National Center for Biotechnology information (NCBI) under BioProject ID ERP001813 and PRJNA260763 and processed these data with the same pipeline. The variation detection followed the best practice workflow recommended by GATK. HaplotypeCaller and GenotypeGVCFs algorithms were jointly used to call variants. Intermediate genomic (gVCF) file were generated using the “-ERC GVCF” mode in “HaplotypeCaller.” Joint genotype was performed using “GenotypeGVCFs” and then subsequently merged with BCFtools (Li et al., 2009). We adopted following hard filtering criteria to retain high-quality variations with “Depth >4.0, MQ RMS mapping quality >20 and -cluster 2, -window 4” for SNPs: “QD <2.0 || FS >200.0 || ReadPosRankSum <-20.0” for insertions/deletions (InDels). After filtering, the variants were annotated with the ANNOVAR software based on gene- and region-based model (Wang et al., 2010).

Population Genetic Structure and Linkage Disequilibrium

To investigate the genetic relationships between ASP and Asian wild pig population, we filtered all autosome single nucleotide variants (SNVs) with a minor allele frequency (MAF) <0.05 and linkage disequilibrium (r^2) <0.2, site missing rate <0.05, and quality value <30 for principal component analysis (PCA), neighbor-joining phylogenetic trees, and linkage disequilibrium; and – indep-pairwise 100 10 0.01 using plink 1.9 for structure. Neighbor-joining (NJ) phylogenetic trees were built based on identical-by-state distance matrix using the PHYLIP v.3.695 package (Felsenstein, 1989). In addition, PCA analyses were performed using the GCTA software (v.1.25) (Yang et al., 2011), and the first three eigenvectors were plotted. Moreover, the population ancestry was inferred by ADMIXTURE (v1.3.0) (Holsinger and Weir, 2009) with a fast-maximum likelihood method. The optimum number of ancestral clusters K was estimated with the five-fold cross-validation procedure. The genome-wide linkage disequilibrium (LD) pattern between ASP and AWB population were assessed using the PopLDdecay software¹ to calculate the average r^2 value with the default parameters.

Genome-Wide Selective Sweeps Detection

Before performing detecting signatures of selection, we filtered the SNVs with call rates <0.90 and MAF <0.05 and remove sites with a missing rate >20% using VCFtools. Even though our Asian wild boars may not fully reflect the genetic diversity of the true progenitor wild Asian pig population, we decided to

¹<https://github.com/BGI-shenzhen/PopLDdecay>

identify some potential selective signals during pig domestication (ASP versus Asian wild boar) by surveying the genomic regions with polymorphism levels ($\theta\pi$, pairwise nucleotide variation as a measure of variability) and genetic differentiation (F_{ST}) based cross approaches, using a 100-kb sliding window approach with 10 kb step size to calculate F_{ST} and $\theta\pi$ values with PopGenome (Pfeifer et al., 2014). The overlapped windows within the top 1% of the F_{ST} and $\theta\pi$ ratio empirical distributions were considered as the candidate selective regions and were subsequently examined for the candidate genes (Li et al., 2013).

Annotation of the Selected Genes and Quantitative Trait Loci Analysis

To further explore the potential biological significance of genes within these sweep regions, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment analyses were carried out through the Database for Annotation, Visualization and Integrated Discovery (DAVID, v.6.8) (Huang et al., 2009). Only terms with a $p < 0.05$ were considered to be significant. Additionally, the pig quantitative trait loci (QTL) database² was used to annotate potential traits related to the potential selection regions based on the physical position of the QTLs.

Protein–Protein Interaction Network Analysis of the Selected Genes

To investigate the interaction associations of selected genes, we applied the selected genes to the Search Tool for the Retrieval of Interacting Genes (STRING, ³) (Szklarczyk et al., 2015), a tool to retrieve and display the genes a query gene repeatedly occurs with in clusters on the genome. The selected genes were mapped and the interactions with Default confidence cutoff of 400 was used. Afterward, a protein–protein interaction (PPI) network was constructed and visualized by Cytoscape software (version 3.4.0,⁴).

RESULTS

Genomic Variant Identification in ASP Breed

To detect genome-wide variation in ASP breed, we performed whole-genome resequencing of 20 unrelated ASPs, which yield 469.01 Gb of sequence data with an average depth of $9\times$ (Table 1). The data were uploaded to the NCBI with BioProject ID (PRJNA634804). After variants calling and subsequent stringent quality filtering, a total of ~ 26 million SNVs with high quality were finally retained, of which 391,337 SNVs were newly identified (not included in the dbSNP database: ⁵). These novel SNVs were expected to be present at lower frequencies or to be specific to the ASP population, accounting for their

TABLE 2 | Annotation of the single nucleotide variants (SNVs).

Variant type	No. of variants
SNV	26,401,669
Intergenic	15,811,056
downstream	153,338
upstream	149,936
5'UTR	55,005
3'UTR	244,339
Splicing site	1,002
Intron	9,772,851
Coding domain	214,142
Synonymous	140,436
Non-synonymous	72,779

lack of previous detection. For all detected SNVs, the average transition/transversion ratio was 2.29, concordant with the previous report (Kerstens et al., 2009).

Further annotation of these SNVs in the ASP population revealed that they were most abundant in intergenic regions (59.89%) and intronic regions (37.02%), followed by downstream (0.58%), upstream (0.57%), untranslated regions (1.14%), and splicing sites; only 0.81% were located in coding sequences. Of the SNPs present in coding regions, 140,436 were synonymous and 72,779 were non-synonymous (Table 2).

Population Genetic Structure and Linkage Disequilibrium

After filtering, there are 201,945 SNVs for structure analyses and 18,859,059 SNVs for PCA, LD, and phylogenetic trees analyses. To assess the phylogenetic relationship among the pig breeds in this study, unrooted phylogenetic tree analyses revealed genetically distinct cluster according to their type (Figure 2A). The branches of the phylogenetic tree were grouped as expected and were consistent with the results of PCA (Figure 2B), thus revealing clustering into two distinct genetic groups. We performed PCA using GCTA, the same type cluster together. The first two PCs explain 13.4 and 6.12% of the total variation, respectively. To further understand the degree of admixture in the population, $K = 2$ was used. As shown in Figure 2C, it can separate all of the ASPs from Asian wild boars. Using the phased genotypes, linkage disequilibrium, in terms of the correlation coefficient (r^2), was calculated for Anqing six-end-white and the Asian wild pig populations. As shown in Figure 3, the LD decay rates were similar between ASP and AWB populations. The faster LD decay was observed in the AWB population, which indicates that artificial selection can facilitate the increase in LD within a population.

Identification of Selective Loci

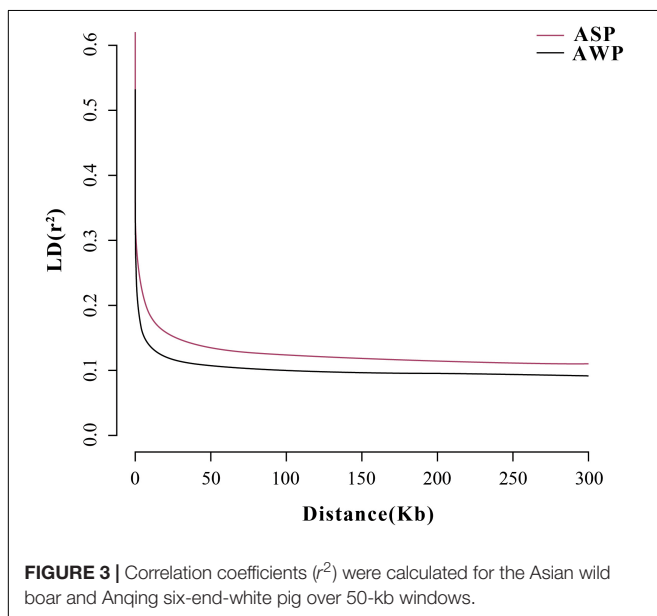
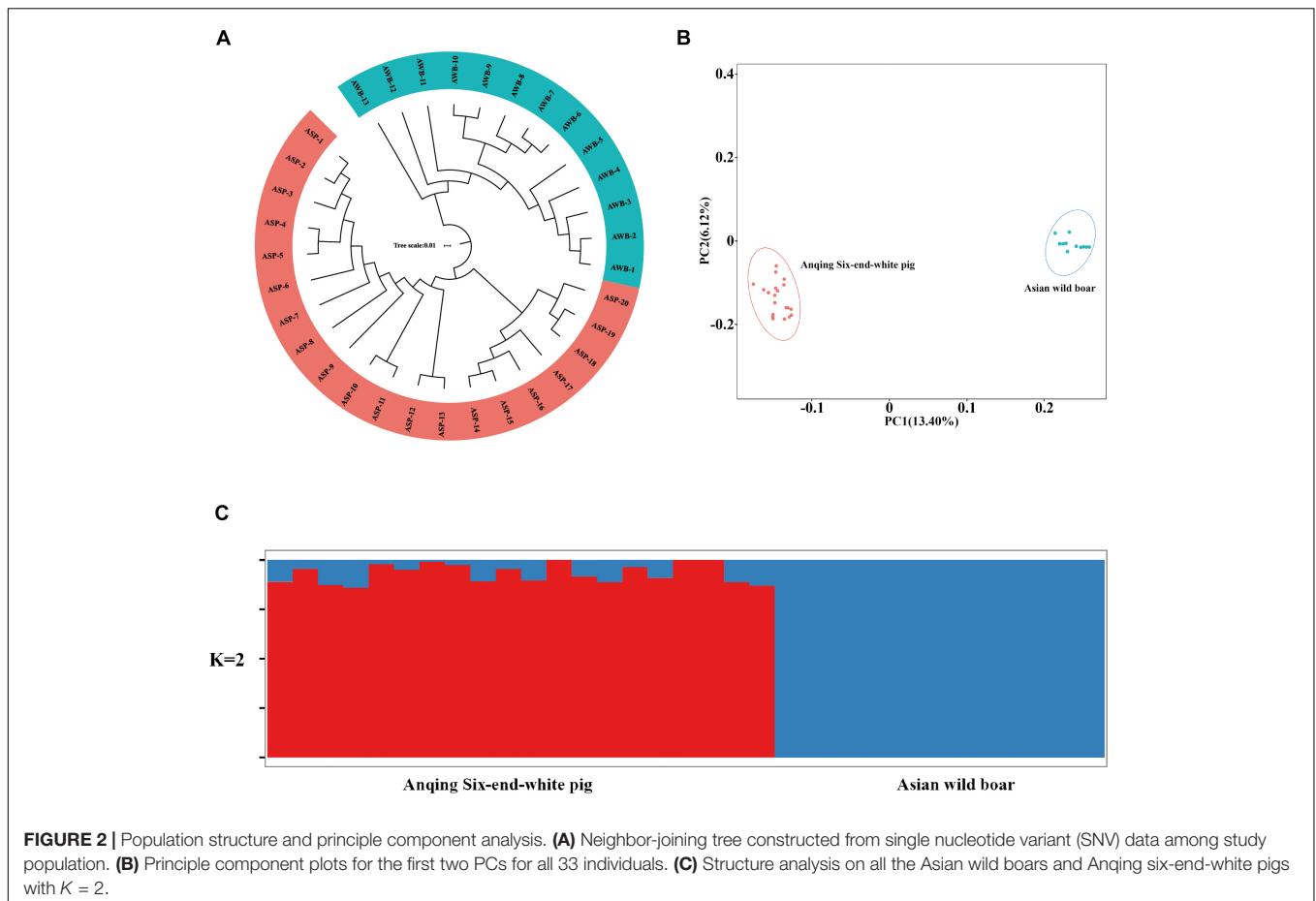
After filtering, there are 21,637,726 SNVs used for signatures of selection detection. We used both the polymorphism levels ($\theta\pi$, pairwise nucleotide variation as a measure of variability) and genetic differentiation (F_{ST}) based cross approaches to investigate the selection signals across the whole genome. In this study, we selected regions both meeting the top 1% threshold as the selected

²www.animalgenome.org

³http://string-db.org/

⁴http://www.cytoscape.org/

⁵ftp://ftp.ensembl.org/pub/release-99/variation/gvf/sus_scrofa/

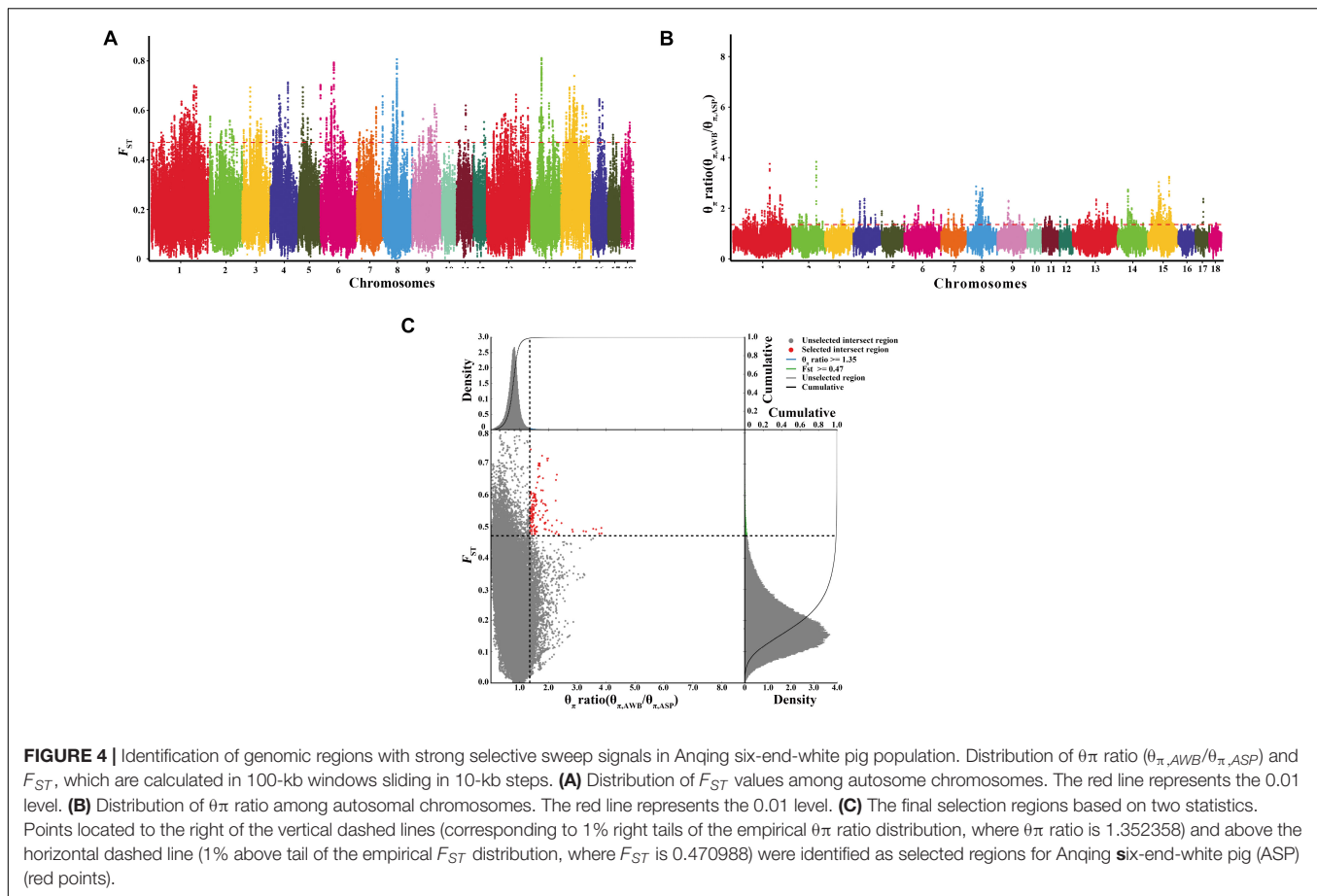


157 selected regions (15.7 Mb of the genome, **Figure 4C** and **Supplementary Table S2**) with extremely high F_{ST} values and significantly high $\theta\pi$ ratios, which meet both top 1% of two values (threshold, 1%; F_{ST} , 0.470988; $\theta\pi$ ratio, 1.352358). On average, there are 70 SNVs in each window. A total of 48 genes harbored in these regions (**Supplementary Table S3**). Further, the SNPs in the selected genes were extracted. There is a total of 25,077 SNVs in the 48 genes, and they were most abundant in intronic regions (94.7%), followed by untranslated regions (2.93%), and only 2.32% were located in coding sequences. Of the SNPs present in coding regions, 338 were synonymous and 245 were non-synonymous (**Supplementary Table S4**). Thirty synonymous were randomly selected and validated by Sanger sequencing. The synonymous and primer information were shown in **Supplementary Tables S5, S6** separately. The results of Sanger sequencing were consistent with the whole-genome resequencing.

Annotation of the Selected Genes and Quantitative Trait Loci Analysis

In order to assess the function of these genes, GO and KEGG analyses were conducted. After enrichment analysis of the selected genes, there are 56 cellular component term, 44 molecular function terms, and 107 biological process terms. In

regions. There are 2,264 selective regions for each approach, which covered 10% of the genome. The genome distribution of the two statistics are shown in **Figures 4A,B**. There are



the GO term level 2, 31 GO terms were enriched (**Supplementary Table S7**). Most of these genes were related to reproduction (7 genes), immune system process (6 genes), growth (3 genes), and response to stimulus (18 genes) (**Figure 5**). In the KEGG analysis, a total of 14 pathways were enriched, and some important pathways were also found, although there is only one gene in the pathway (**Supplementary Table S8**). Most of pathways were related to carbohydrate digestion and absorption (three genes), salivary secretion (two genes), olfactory transduction (one gene), peroxisome proliferator-activated receptor (PPAR) signaling pathway (one gene), and MAPK signaling pathway (one gene) (**Figure 6**). We also found a selected gene (MC1R), which could explain the coat color phenotype of ASP. QTL overlapping with the potential selection regions detected by two methods was associated with meat and carcass (45.97%), reproduction association (11.09%), production (10.42%), and so on, as shown in **Supplementary Table S9**.

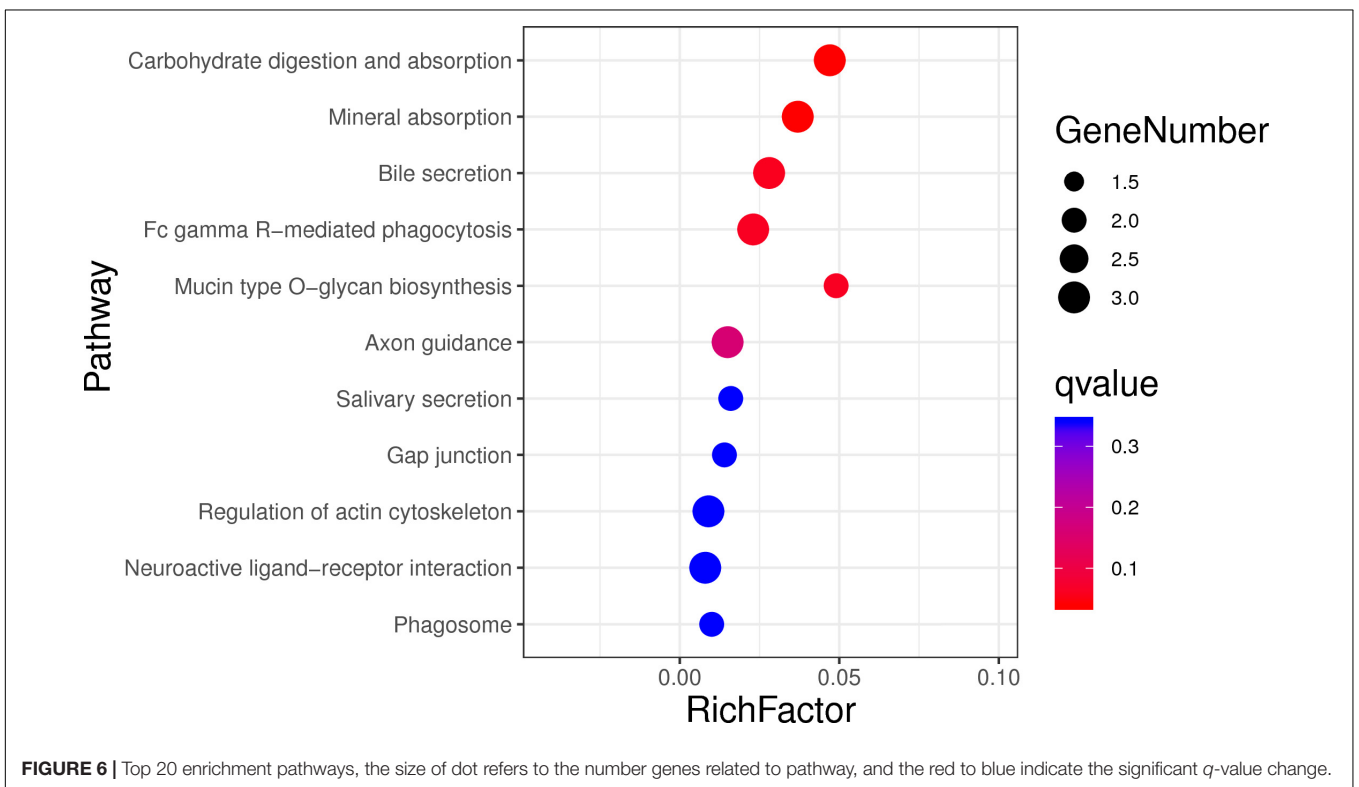
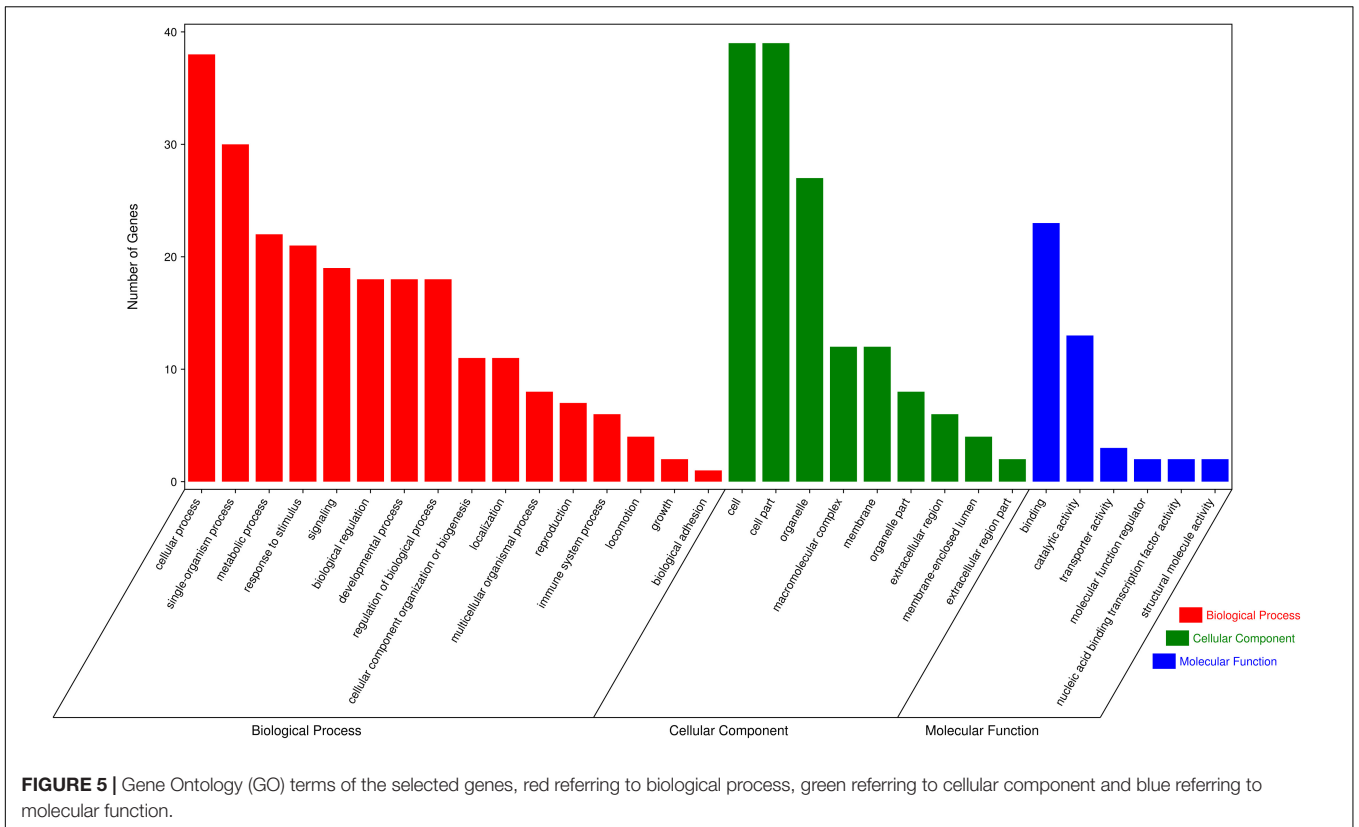
PPI Network Construction

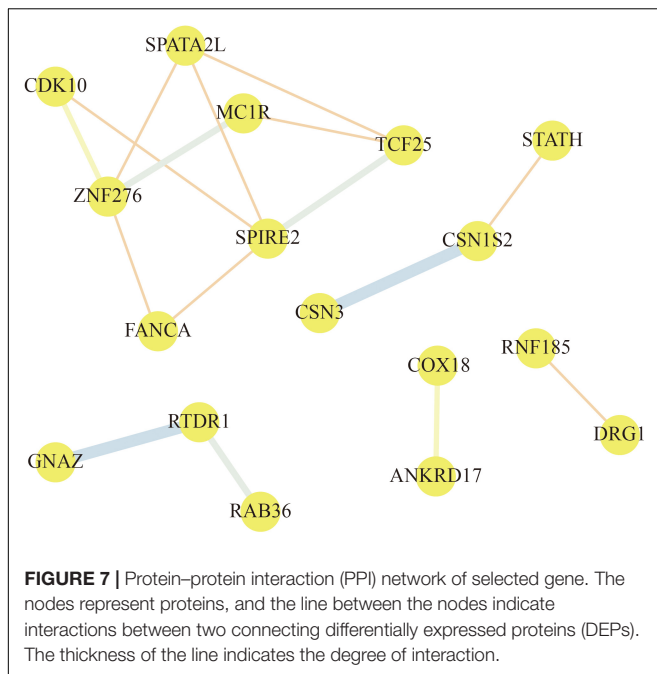
To explore the relationships among these genes, a PPI network of the selected genes was constructed by STRING and visualized by Cytoscape. According to the node pair combing score ≥ 0.4 , a total of 17 selected genes were filtered into the PPI network that interacted with other selected genes (**Figure 7**). As shown in **Figure 7**, the PPI network of selected genes comprised 17 nodes

and 16 edges. From this analysis of selected genes, we focused on the selected genes that interacted with three or more other genes. ZNF276, SPIRE2, TCF25, and SPATA2L are found to be the hub genes in the network. Based on these results, we assumed that ZNF276, SPIRE2, TCF25, and SPATA2L could be promising candidate genes that affect “cellular process,” “metabolic process,” “cell,” “macromolecular complex,” etc., indicating that these genes might play an important role in metabolism and cell process.

DISCUSSION

As one of the first domesticated animals, pig has played an important role in many aspects of human life. The ancestors of domesticated pig present in the world provide a unique opportunity for elucidating the genetics basis of domestication and further promoting the breeding of pig. In recent years, selection signatures have been identified in agricultural animals, leading to the elucidation of the mechanisms of many complex traits. To better understand the genetic basis underlying domestication and natural selection, we performed whole-genome resequencing on 20 unrelated ASP combined with 13 downloaded Asian wild boars. We selected windows with simultaneously high F_{ST} values (1% right tail) and significantly high $\theta\pi$ ratios (1% right tail). There are 157 selected regions





with 48 genes harbored in these regions. Functional enrichment analyses revealed that the selected genes may play an important role in reproduction, immune system process, growth, salivary secretion, coat color, and other traits.

We found that a gene could elucidate the genetic basis of coat color phenotype of ASP population. In previous studies about pig coat color, *MC1R* gene was under positive selection (Li et al., 2010; Zhao et al., 2018). In this study, the *MC1R* gene was also found under selection based on cross approaches. To determine whether the six-end-white phenotype in the ASP is associated with the 2-bp insertion in the coding sequence of the *MC1R* gene, as previously shown in Bama miniature pigs (Jia et al., 2017), we searched for the presence of a 2-bp insertion in *MC1R* not detected in ASP population. This observation implies that the six-end-white phenotype in the ASP is not caused by a 2-bp insertion in the *MC1R* gene. We found five missense variants (exon1: c. G283A: p.V95M; exon1: c.T305C: p. L102P; exon1: c. A727G: p. T243A; exon1: c. T491C: p.V164A; exon1: c. G364A: p.V122I) within the *MC1R* gene. The amino acids changed by G283A and T503C were identified in ASP but absent from wild boars, which is consistent with a previous study (Li et al., 2010). Meanwhile, the “exon1: c.T305C: p. L102P” variant could result in a Leu to Pro substitution. Evidence from other species strongly suggest that the change from a Leu at codon to a Pro substitution was associated with a dominant black color trait (Klungland et al., 1995; Kijas et al., 1998). These results further imply that the ASP coat color is likely attributed by *MC1R* mutations. Therefore, the identification in *MC1R* helps us to better explain the black coat of the ASP breed. Besides these well-known loci, further investigation about the effect of other missense variants on coat color phenotypic variation in ASP needs to be done. The putative role of these variants should be tested with functional experiments.

We also found a selected gene related to “salivary secretion.” Saliva played an important role as lubricant and an antimicrobial, preventing the dissolution of teeth, aiding in digestion, and facilitating taste (Carpenter, 2013). In a previous study, it has been proven that domestic pig produces more saliva than Tibetan wild boars and found that *KCNMA1* and *TRPC1* exhibited strong selective signals (Li et al., 2013). Even though the two genes were not harbored in the selection regions, we also found a gene Statherin (*STATH*) related to “salivary secretion” pathway, which has been under selection pressure in ASP. Statherin is a salivary protein encoded by the *STATH* gene, which helps to control the formation of hydroxyapatite crystals and plays an important role in maintaining the tooth enamel in the oral cavity (Schwartz et al., 1992; Goobes et al., 2006). The *STATH* gene is expressed in the human, dog, and pig salivary gland (Schlesinger and Hay, 1977; Nakanishi et al., 2016), and it was confirmed that the messenger RNA (mRNA) of the *Statherin* (*STATH*) gene was present in saliva but absent in blood, semen, vaginal secretions, and menstrual blood (Tsai et al., 2018). Positive selection of the *STATH* gene is consistent with the observation that domestic pig produces more saliva than wild boars (Li et al., 2013).

As is well known, the domestic pigs have a higher fertility than wild boars. We identified several genes involved in GO term related to reproduction. They played an important role in asymmetry, spermatogenesis, embryo cleavage and blastocyst formation, meiosis and germ cell development, and acrosome reaction. Of these genes, spire type actin nucleation factor 2 (*SPIRE2*) was found playing a key factor in asymmetric division of mouse oocytes, and the mRNA levels of *SPIRE2* in oocytes are significantly higher than in other tissues (Pfinder et al., 2011). Meanwhile, the asymmetric oocyte division is essential for fertility (Leader et al., 2002). In previous studies, oxysterol-binding protein 2 (*OSBP2*) has been elucidated playing an important role in the postmeiotic differentiation of germ cells and would cause male infertility owing to oligoastheno-teratozoospermia with lack of *OSBP2* (*ORP4*) (Charman et al., 2014; Udagawa et al., 2014). This may imply that *OSBP2* is significantly associated with spermatogenesis. *LIM kinases 2* (*LIMK2*), especially the testis-specific isoform tLIMK2, is specifically expressed in differentiated, meiotic stages of spermatogenic cells and plays an important role in proper progression of spermatogenesis by the regulation of cofilin activity and/or localization in germ cells (Takahashi et al., 2002). The weight of the testes in *LIMK2*–/– mice was significantly reduced to approximately 80% compared to that of control mice (Takahashi et al., 2002). In addition, the inhibition of *LIMK1/2* activity in mouse causes the failure of embryo cleavage and blastocyst formation (Duan et al., 2018). Fanconi anemia (*FANCA*) genes, traditionally known for their essential roles in DNA repair and cytogenetic instability, have been demonstrated to be involved in meiosis and germ cell development. In a previous study about premature ovarian insufficiency (POI), two missense variants of *FANCA* were identified and could reduce its protein expression level compared with non-POI women. Meanwhile, heterozygous mutated female mice (*Fanca*±) showed reduced fertility and declined numbers of follicles with aging when compared with the wild-type female

mice (Yang et al., 2019). Calcium-binding protein, sperm-specific 1 (*CABS1*) was first reported as one of the genes highly expressed in human testis (Jiahao et al., 2000) and specifically expressed in mice in the elongate spermatids and then localized into the principal piece of flagella of matured spermatozoa (Kawashima et al., 2009). Shawki et al. found that the porcine *CABS1* localizes to the acrosome in addition to the tail where m*CABS1* only localizes in mature sperm, suggesting that porcine *CABS1* is involved in the acrosome reaction (Shawki et al., 2016).

In a previous study, it has been reported that the larger and more diverse olfactory gene repertoire may help pig to recognize odors that disseminated from a wide range of food types and flavoring agents that are present in artificial feeds, which result in higher feed intake and pork yield in duroc pig (Nguyen et al., 2012). It is well known that the feed intake and pork yield of duroc are higher than that of Chinese domesticated pigs. On the other hand, when comparing to wild boars, the Chinese domesticated pigs had a higher feed intake and pork yield, which may result from the olfactory genes in domesticated pigs. In this study, a gene (*OR5K4*) in ASP was selected and related to “olfactory transduction,” we thus assume that selection on specific olfactory receptors could enable domestic pigs to have a higher feed intake and pork yield. This may explain the signatures of selection at the olfactory receptor gene and the Chinese domesticated pigs having a higher feed intake and pork yield than wild boars.

Several genes related to “growth” were also identified. Neuropeptide Y (NPY) is widely expressed in the central nervous system and influences many physiological processes, including food intake (Zarjevski et al., 1993) and bone density (Teixeira et al., 2009). Meanwhile, Y1 and Y5 receptors (*NPY1R* and *NPY5R*) are expressed in hypothalamic areas that control feeding (Larsen et al., 1993; Parker and Herzog, 1999). It has been demonstrated that both *NPY1R* and *NPY5R* played a key role in the control of food intake in mice (Raposo et al., 2004). Selenoprotein M (*SELENOM*), a positive regulator of leptin signaling and thioredoxin antioxidant activity in the hypothalamus, has been demonstrated to play a key role in Ca^{2+} homeostasis and energy metabolism (Gong et al., 2019). Mice with *Selenom*^{-/-} knockout exerted a significant influence on energy homeostasis, including an increased body weight and reduced hypothalamic leptin sensitivity (Pitts et al., 2013).

Although some interesting findings were reported here, the limitations of the present study should not be neglected. On the one hand, the alleles were described from a limited sample size of ASP ($n = 20$) and wild boars ($n = 13$), which might not completely represent the populations and affect the F_{ST} and $\theta\pi$ ratio statistic. On the other hand, the function of these selected genes was annotated with GO and KEGG database, and although studies have been done, there is still a need to conduct investigations to understand the underlying genetic mechanism on traits in pig. The limitations might impact the observations of this study and should be overcome in further investigations.

This study detected the genomic signatures of selection that may have shaped the domestication of ASP in China. The genes found to be positively selected in ASP are involved

in crucial biological processes such as coat color (*MC1R*), salivary secretion (*STATH*), reproduction (*SPIRE2*, *OSBP2*, *LIMK2*, *FANCA*, and *CABS1*), olfactory transduction (*OR5K4*), and growth (*NPY1R*, *NPY5R*, and *SELENOM*). In addition, mutations within these genes were also identified, which can be used to further refine selection in ASP in the future. Our research increased the knowledge of ASP phenotype-related genes and helped to improve our understanding of the underlying biological mechanisms.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study reviewed and approved in this study was carried out in accordance with the recommendations of the Animal Care Committee of Anhui Academy of Agricultural Sciences (Hefei, China). The protocol was approved by the Animal Care Committee of Anhui Academy of Agricultural Sciences (No. AAAS2020-04).

AUTHOR CONTRIBUTIONS

WZ, CW, and ZY: conceptualization. WZ, MZ, MY, YW, XW, XZ, and YD: data curation. WZ: writing—original draft preparation and project administration. WZ, CW, ZY, and GZ: writing—review and editing. CW and ZY: funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the National Natural Science Foundation of China (Nos. 31572377, 31972531, and 31371258), the planning subject of “the 13th 5-Year-Plan” (No. 2017YFD0600805), the Anhui Academy of Agricultural Sciences Key Laboratory Project (Nos. 2020YL031 and 17030701008), Anhui Science and Technology Key Project (2008085MC87), Anhui Swine Industry Technology System Project (No. AHCYTX-05-09), and Anhui Pig Genetic Improvement Project (No. 2018FACN4373).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.566255/full#supplementary-material>

REFERENCES

- Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., et al. (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* 47, 217–225. doi: 10.1038/ng.3199
- Akey, J. M., Zhang, G., Zhang, K., Jin, L., and Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12, 1805–1814. doi: 10.1101/gr.631202
- Carpenter, G. H. (2013). The secretion, components, and properties of saliva. *Ann. Rev. Food Sci. Technol.* 4, 267–276. doi: 10.1146/annurev-food-030212-182700
- Charman, M., Colbourne, T. R., Pietrangolo, A., Kreplak, L., and Ridgway, N. D. (2014). Oxysterol-binding protein (OSBP)-related protein 4 (ORP4) is essential for cell proliferation and survival. *J. Biol. Chem.* 289, 15705–15717. doi: 10.1074/jbc.M114.571216
- Chen, C., Liu, C., Xiong, X., Fang, S., Yang, H., Zhang, Z., et al. (2018). Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet. Select. Evol.* 50:72. doi: 10.1186/s12711-018-0442-6
- Ding, Y., Qian, L., Wang, L., Wu, C., Li, D., Zhang, X., et al. (2020). Relationship among porcine lncRNA TCONS_00010987, miR-323, and leptin receptor based on dual luciferase reporter gene assays and expression patterns. *Asian Austr. J. Anim. Sci.* 33, 219–229. doi: 10.5713/ajas.19.0065
- Duan, X., Zhang, H. L., Wu, L. L., Liu, M. Y., Pan, M. H., Ou, X. H., et al. (2018). Involvement of LIMK1/2 in actin assembly during mouse embryo development. *Cell Cycle* 17, 1381–1389. doi: 10.1080/15384101.2018.1482138
- Fan, H., Wu, Y., Qi, X., Zhang, J., Li, J., Gao, X., et al. (2014). Genome-wide detection of selective signatures in simmental cattle. *J. Appl. Genet.* 55, 343–351. doi: 10.1007/s13353-014-0200-6
- Fang, M., Larson, G., Ribeiro, H. S., Li, N., and Andersson, L. (2009). Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* 5:e1000341. doi: 10.1371/journal.pgen.1000341
- Felsenstein, J. (1989). PHYLIP - Phylogeny inference package (version 3.2). *Cladistics* 5, 164–166.
- Frantz, L. A., Schraiber, J. G., Madsen, O., Megens, H. J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47, 1141–1148. doi: 10.1038/ng.3394
- Gianola, D., Simianer, H., and Qanbari, S. (2010). A two-step method for detecting selection signatures using genetic markers. *Genet. Res.* 92, 141–155. doi: 10.1017/S0016672310000121
- Giuffra, E., Kijas, J. M., Amarger, V., Carlborg, O., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785–1791.
- Gong, T., Hashimoto, A. C., Sasuclark, A. R., Khadka, V. S., Gurary, A., and Pitts, M. W. (2019). Selenoprotein M promotes hypothalamic leptin signaling and thioredoxin antioxidant activity. *Antioxid. Redox Signal.* doi: 10.1089/ars.2018.7594 [Epub ahead of print].
- Goobes, R., Goobes, G., Campbell, C. T., and Stayton, P. S. (2006). Thermodynamics of statherin adsorption onto hydroxyapatite. *Biochemistry* 45, 5576–5586. doi: 10.1021/bi052321z
- Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622
- Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* 10, 639–650. doi: 10.1038/nrg2611
- Hu, H., Wu, C., Ding, Y., Zhang, X., Yang, M., Wen, A., et al. (2019). Comparative analysis of meat sensory quality, antioxidant status, growth hormone and orexin between Anqingliubai and Yorkshire pigs. *J. Appl. Anim. Res.* 47, 357–361. doi: 10.1080/09712119.2019.1643729
- Huang, D., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Jia, Q., Cao, C., Tang, H., Zhang, Y., Zheng, Q., Wang, X., et al. (2017). A 2-bp insertion (c.67_68insCC) in MC1R causes recessive white coat color in Bama miniature pigs. *J. Genet. Genom.* 44, 215–217. doi: 10.1016/j.jgg.2017.02.003
- Jiahao, S., Zuomin, H., and Jianmin, L. (2000). “Preparation of human testicular cDNA microarray and initial research of gene expression library related to spermatogenesis,” in *Epithelial Cell Biology—A Primer*, ed. H. S. Chan (Beijing: Press of Military Medical Science), 274–277.
- Kawashima, A., Osman, B. A., Takashima, M., Kikuchi, A., Kohchi, S., Satoh, E., et al. (2009). CABS1 is a novel calcium-binding protein specifically expressed in elongate spermatids of mice. *Biol. Reprod.* 80, 1293–1304. doi: 10.1095/biolreprod.108.073866
- Kerstens, H. H., Kollers, S., Kommadath, A., Del Rosario, M., Dibbits, B., Kinders, S. M., et al. (2009). Mining for single nucleotide polymorphisms in pig genome sequence data. *BMC Genom.* 10:4. doi: 10.1186/1471-2164-10-4
- Kijas, J. M., Wales, R., Törnsten, A., Chardon, P., Moller, M., and Andersson, L. (1998). Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* 150, 1177–1185.
- Kijas, J. M. H., and Andersson, L. (2001). A Phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *J. Mol. Evol.* 52, 302–308. doi: 10.1007/s002390010158
- Klungland, H., Våge, D. I., Gomez-Raya, L., Adalsteinsson, S., and Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm. Genome* 6, 636–639. doi: 10.1007/BF00352371
- Larsen, P. J., Sheikh, S. P., Jakobsen, C. R., Schwartz, T. W., and Mikkelsen, J. D. (1993). Regional distribution of putative NPY Y1 receptors and neurons expressing Y1 mRNA in forebrain areas of the rat central nervous system. *Eur. J. Neurosci.* 5, 1622–1637. doi: 10.1111/j.1460-9568
- Larson, G. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307, 1618–1621. doi: 10.1126/science.1106927
- Leader, B., Lim, H., Carabatsos, M. J., Harrington, A., Ecsedy, J., Pellman, D., et al. (2002). Formin-2, polyploidy, hypofertility and positioning of the meiotic spindle in mouse oocytes. *Nat. Cell Biol.* 4, 921–928. doi: 10.1038/ncb880
- Lewontin, R. C., and Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74, 175–195.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, J., Yang, H., Li, J. R., Li, H. P., Ning, T., Pan, X. R., et al. (2010). Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. *Heredity* 105, 274–281. doi: 10.1038/hdy.2009.191
- Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi: 10.1038/ng.2811
- Nakanishi, H., Ohmori, T., Hara, M., Yoneyama, K., Takada, A., and Saito, K. (2016). Identification of canine saliva using mRNA-based assay. *Intern. J. Legal Med.* 131, 39–43. doi: 10.1007/s00414-016-1391-7
- Nei, M., and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 76, 5269–5273. doi: 10.1073/pnas.76.10.5269
- Nguyen, D. T., Lee, K., Choi, H., Choi, M. K., Le, M. T., Song, N., et al. (2012). The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome. *BMC Genom.* 13:584. doi: 10.1186/1471-2164-13-584
- Parker, R. M., and Herzog, H. (1999). Regional distribution of Y-receptor subtype mRNAs in rat brain. *Eur. J. Neurosci.* 11, 1431–1448. doi: 10.1046/j.1460-9568.1999.00553.x
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31, 1929–1936. doi: 10.1093/molbev/msu136
- Pfender, S., Kuznetsov, V., Pleiser, S., Kerkhoff, E., and Schuh, M. (2011). Spire-type actin nucleators cooperate with Formin-2 to drive asymmetric oocyte division. *Curr. Biol. CB* 21, 955–960. doi: 10.1016/j.cub.2011.04.029
- Pitts, M. W., Reeves, M. A., Hashimoto, A. C., Ogawa, A., Kremer, P., Seale, L. A., et al. (2013). Deletion of selenoprotein M leads to obesity without cognitive deficits. *J. Biol. Chem.* 288, 26121–26134. doi: 10.1074/jbc.M113.471235
- Raposo, P. D., Pedrazzini, T., White, R. B., Palmiter, R. D., and Aubert, M. L. (2004). Chronic neuropeptide Y infusion into the lateral ventricle induces

- sustained feeding and obesity in mice lacking either Npy1r or Npy5r expression. *Endocrinology* 145, 304–310. doi: 10.1210/en.2003-0914
- Rubin, C. J., Megens, H. J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19529–19536. doi: 10.1073/pnas.1217149109
- Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd Edn, New York, NY: Cold Spring Harbor Laboratory Press.
- Schlesinger, D. H., and Hay, D. I. (1977). Complete covalent structure of statherin, a tyrosine-rich acidic peptide which inhibits calcium phosphate precipitation from human parotid saliva. *J. Biol. Chem.* 252, 1689–1695.
- Schwartz, S. S., Hay, D. I., and Schluckebier, S. K. (1992). Inhibition of calcium phosphate precipitation by human salivary statherin: structure-activity relationships. *Calcif. Tissue Intern.* 50, 511–517. doi: 10.1007/BF00582164
- Shawki, H. H., Kigoshi, T., Katoh, Y., Matsuda, M., Ugboma, C. M., Takahashi, S., et al. (2016). Identification, localization, and functional analysis of the homologues of mouse CABS1 protein in porcine testis. *Exper. Anim.* 65, 253–265. doi: 10.1538/expanim.15-0104
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, 447–D452. doi: 10.1093/nar/gku1003
- Takahashi, H., Koshimizu, U., Miyazaki, J., and Nakamura, T. (2002). Impaired spermatogenic ability of testicular germ cells in mice deficient in the LIM-kinase 2 gene. *Dev. Biol.* 241, 259–272. doi: 10.1006/dbio.2001.0512
- Teixeira, L., Sousa, D. M., Nunes, A. F., Sousa, M. M., Herzog, H., and Lamghari, M. (2009). NPY revealed as a critical 605 modulator of osteoblast function in vitro: new insights into the role of Y1 and Y2 receptors. *J. Cell. Biochem.* 107, 908–916. doi: 10.1002/jcb.22194
- Tsai, L. C., Su, C. W., Lee, J. C., Lu, Y. S., Chen, H. C., Lin, Y. C., et al. (2018). The detection and identification of saliva in forensic samples by RT-LAMP. *Foren. Sci. Med. Pathol.* 14, 469–477. doi: 10.1007/s12024-018-0008-5
- Udagawa, O., Ito, C., Ogonuki, N., Sato, H., Lee, S., Tripvanuntakul, P., et al. (2014). Oligo-astheno-teratozoospermia in mice lacking ORP4, a sterol-binding protein in the OSBP-related protein family. *Genes Cells* 19, 13–27. doi: 10.1111/gtc.12105
- Veirano Fréchou, R. (2007). The state of the world's animal genetic resources for food and agriculture. *Acta Paediatr.* 81, 21–24.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603
- Wang, Y., Zhang, W., Wu, X., Wu, C., Qian, L., Wang, L., et al. (2020). Transcriptomic comparison of liver tissue between Anqing six-end-white pigs and Yorkshire pigs based on RNA sequencing. *Genome* 63, 203–214. doi: 10.1139/gen-2019-0105
- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 1975, 256–276. doi: 10.1016/0040-5809(75)90020-9
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646
- Wright, S. (1949). The genetical structure of populations. *Ann. Eugen.* 15, 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, X., Zhang, X., Jiao, J., Zhang, F., Pan, Y., Wang, Q., et al. (2019). Rare variants in FANCA induce premature ovarian insufficiency. *Hum. Genet.* 138, 1227–1236. doi: 10.1007/s00439-019-02059-9
- Zarjevski, N., Cusin, I., Vettor, R., Rohner-Jeanrenaud, F., and Jeanrenaud, B. (1993). Chronic intracerebroventricular neuropeptide-Y administration to normal rats mimics hormonal and metabolic changes of obesity. *Endocrinology* 133, 1753–1758. doi: 10.1210/endo.133.4.8404618
- Zhang, X. D., Zhang, S. J., Ding, Y. Y., Feng, Y. F., Zhu, H. Y., Huang, L., et al. (2015). Association between ADSL, GARS-AIRS-GART, DGAT1, and DECR1 expression levels and pork meat quality traits. *Genet. Mol. Res.* 14, 14823–14830. doi: 10.4238/2015.November.18.47
- Zhao, P., Yu, Y., Feng, W., Du, H., Yu, J., Kang, H., et al. (2018). Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *Gigascience* 7:giy058. doi: 10.1093/gigascience/giy058
- Zhu, J., Chen, C., Yang, B., Guo, Y., Ai, H., Ren, J., et al. (2015). A systems genetics study of swine illustrates mechanisms underlying human phenotypic traits. *BMC Genom.* 16:88. doi: 10.1186/s12864-015-1240-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhang, Yang, Zhou, Wang, Wu, Zhang, Ding, Zhao, Yin and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.